



Appendix A

BsgI restriction endonuclease (RM fusion) coding sequence
3108 nt (recognition sequence: GTGCAG)

ATGAATAGAG TAGAATCTAA AAAAAAATTA GAACAATTAG TTCAACAGTT CGAGAAGTAT
GAAAGTACAT ATAGCGCTTC GGATTATAAA GAGGCAACTT TAAGATCTAG TTTTTAGAT
CCCTTTTG AACTTTGG ATGGAAATG CGCCCTGAAA GAATAACTAA TCCAGCAGAC
TTAGAAGTGA TTATAGAAGA AAGTTAGAA ACGGAAAAAT CTACTAAGTA TATAGATTAT
GTTTTAAAAA TTAATAGAAC GACTCAGTTT TTGGTAGAAG CTAAAAGCC AGCTGAAAGT
CTTTCTAAA AAGATCATAT TTTTCAGGCT AAAAGTTATG CATTACTAC GGAGATTCCA
TTTGTCAATT TAACAAATT TAAAGAGTTC AGATTTATG ACGTTCAAC TGAACCTTA
CACAATCAAC CGGATACAGA TAAAGTGGAA GAATATTGTT TTGATTATAA AGAATATGTT
CAAAACTTG ATAAGTTATG GGAATTATTC AGCAGAGAAG CAGTGCTAA CAGAAGTTA
GCCAAGTTT ATGCTAAAAG AAGAAATATA GTAGATAGTC CAGATTAAAT TTTTAAACTT
AATTATCAA TTGATAAAGG TGCATCATT CTGGATATAT CTTCTTAAA AAATTTGAAA
ATATGGAGAA AATCATTAGC TGAAAATATC TTTAATAATA ATTCACTTAA TGTTAACGTA
ATTAATGAAG TAGTCAGAG AATATTAGAT AGACTGATAT TTATCCGTAT CATTGAAGAC
AGAAATATTG AATCTAAAAGA GTTTTAAAAA GAAATTGTA AGATGCCAGA ACAAGATAAT
TCGATTCAG TGAAAAATGA ACTAGATAAA CTATGTATTG AATTAAATAA GAAATTAAAT
GGGTTAGTTT TCCATGACCA CACATTGTT AACGAAGCGT TGATAGATAA CGAAATTAA
ATAGTTATTA TTGACAATT ATATTATCCA AAGTCTCCTT ATAACTTTAG ATTAATTAAG
CCAGAAAATT TAGGACGAAT ATTCGAGCAA TTTTGGGTG AAAAAATTGA AATAATCGAT
GGAAAATAA CGTTCAGGATT AAAAGATATT AATAAAAAAT CAGGAGGTGT TTATTATACA
CCTTCATATA TAGTTGAAAA AATAGTAGAA AATACATTAT CCAAAATT ACATAATGAT
ATTACTATTG AAAATTAGA ACAGATAAAA ATAGCTGACA TAGCTTGTGG TTCAGGAAGC
TTTTAATT CATCATATAA ATATTAATT GATAAATTC AATATATTAA TTCCAATGT
TCGGAAGCGG ATGTTCAAAC ATTAATTAGT AATAACTTAG TATTATAGA CAATGGTAAA
TTAATGTTAA CAATGGAACA TAAAAGGGG ATACTTCAGC AAAATTTTT TGGGTTAGAT
ATAGATTCA CAGCAATTCA GGTAGCGAAA TTAAGTCTT ATATAACCAT GTTAAAGAA
GGATACAGAG AAGGTACATT AAGACCTATA TTACCAAGACT TAAATGATAA TATTAAACAT
GGTAACTCAA TAATAGATAA TGAAATTAA TTTGAAGATG ATATAAATTA CGATATTGAT
GCAACATTAC CATTGATTG GGAATATGCT TTTCTGATA TTATAGATAA CGGAGGTTT
GATGTAATAT TAGGCAATCC ACCCTATATA AGAATTCAA TTTTGAAGA GTTATATGGA
AAAGATGTAG TTAATTATTG GAAAAAAA TACGTTCTG CCGAAAAATT TAACTTGAT
ATATATGTCG TGTTTATAGA AAAAGCATTG TCACTCTGA ATGACCAGGG GATATTGGGA
TATATTGTA TGAACAAATT TTTACTACA CAATATGGAG AAAATTGCG CGAGTTAATA
ACTTCACAAA AATTATTATA TGAAATCATT GATTTGAA TTAATGAAAT ATTTAATAAT
GCTACTACTT ATACTTGTAT ATTAATTAA GACAAACTA ATCCAGATGA AATAATTATT
GAAAGAGTGA TTGATTAA TACTTGGAAA GCTGGAGAAT CTTCACTG GAAAGTGGTA
GATCATACCG AATTCACTAG TACTCCTGG TATTTATCAA GCAATACTGA TGAAGAAATT
TACAAATTCT TTGAAGAAAA TATGGTTTA CTTGAAACCA TTAGTGTAG GGTGTTGTT
GGTGGTCAGA CAGACTGTGA TCCAGTATAT ATTTAGAAG AAGTTATGA AGAAGAAAAT
TATTATATT GTAAGTCAGA ATATACTACT GAAGTACACA AGTTGAAA AGATCATTAA
AAACCATTAA TAAAAGGTT TCTAGATATA AAGAAATATA CTTTTCAA TGTTAATAAG
TGGTTACTTT TCCCTTATAC CAATTGGAA AATACTCTG ATTAATTCC CGAAACAACT
TACAAACAGT ATTTCCCAGA AACATGGAA TACTTAGAGT CTTGTAAAGA AAGATTAGCA
AAAAGAAAAA GTATTGAAAG AGAATTGGAT ATTAATCCGA ATTATAATGA GTGGTATAAA
TATATTACAA AAAAGAATCA CACGAGGATG GACCAATTAA AAATAGTATT TCCTGCGATA
TCGAAGGGTA GTAGCTTTG TTATGATTG GATGGAGAGT ACTATTTGT AGGAAGTGGT
GCTGGAGGCG GTGGTGGAGG CGCAATAGTC TTGCCAGATC AATCTGATTA TAATTATTTA
TCCTTACTTG GAATTCTAA TTCAGAAGTA GTTCACTATC AAATTGTAAG AAGAGGTTCA
AAACATAAAAG GTTCTTATTA TGGTGTAGAT AAAAGAGAA TAGAAAATCT ATATGTGCCA
TTGATTAATG AGGATAATAA AAATTATTT AGTAATATT CAAAATGGT AGCTCAAATT
CTTGATGCGT TTCAAAAAT GCATCAAGCA GGGACAACGG ATGTTGGTAA AGAACAACTT
CAACAAAGAA TAAAATGCT TAATGCTAGA ATAAATGAGC TGGTATATAG ACTGTATAAT
TTACCAAGTAG AATATAAAGA ATATATTAA AATGCCCTAG AAAATTAA